UNIVERSITI TEKNOLOGI MARA

OBESITY PREDICTIVE PROCESS FRAMEWORK BASED ON DIETARY PATTERNS FROM GROCERY DATA

NUR'AINA BINTI DAUD

Thesis submitted in fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Information Technology)

Faculty of Computer and Mathematical Sciences

April 2021

AUTHOR'S DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Teknologi MARA. It is original and is the results of my own work, unless otherwise indicated or acknowledged as referenced work. This thesis has not been submitted to any other academic institution or non-academic institution for any degree or qualification.

I, hereby, acknowledge that I have been supplied with the Academic Rules and Regulations for Post Graduate, Universiti Teknologi MARA, regulating the conduct of my study and research.

Name of Student	:	Nur'Aina Binti Daud
Student I.D. No.	:	2014208596
Programme	:	Doctor of Philosophy (Information Technology) – CS951
Faculty	:	Computer and Mathematical Sciences
Thesis Title	:	Obesity Predictive Process Framework Based on Dietary Patterns from Grocery Data
Signature of Student	:	……..
Date	:	April 2021

ABSTRACT

Globally, obese population prevalence is estimated using Body Mass Index (BMI) data from collected surveys and generally analyzed using simple linear regression method. Even the World Health Organization (WHO) uses this simple BMI index to define overweight and obesity. However, this BMI method of estimating national obesity population prevalence, although simple, is excessively costly involving huge amount of public health data. The aim of the study is to provide an alternatively cheaper obesity predictive analytics using dietary patterns to overcome the current exorbitantly high operation cost. Exploring the feasibility of using grocery data, these data are assessed from physical receipts collected from the selected household sampling which is generated into nutrition data, and then stored in a food engine (G2NE) developed in this study. Data mining technology is then applied on the tested nutrition data to run the obesity predictive analytics. In this study, BMI data are used as a complementary variable to nutrition variable in estimating the individual nutrition consumption, which are assigned as input data in testing. The prediction modelling is developed by conducting the processes in exploratory phase. The findings from the predictive analytics process in exploratory phase show that the use of 70:30 training and testing data split are the best test options for nutrition dataset based on Area Under the Curve (AUC) performance measurement. From the analysis, obese household shows the patterns that have higher intake percentage of processed food of food pyramid level three compared to non-obese household. The processes in the proposed prediction method involve three different domains which are retail, nutrition and health. In bridging the process involving these three domains, the obesity prediction method in this study is proposed in a form of a process framework (G2NOP Framework). This study has come out with two types of research contributions which are practical contributions from data pre-process phase as well as methodological contributions from the exploratory phase and framework integration phase. The limitations in this study include the nutrition assumptions and the scope of the respondents. Hence, the continuation of the future work is to extend the G2NE with the implementation of grocery and nutrition data sharing from the grocery retailers in which the use of larger scope of the household sampling should be considered.

ACKNOWLEDGEMENT

Praise to ALLAH s.w.t. for HIS grace, mercy and blessing bestowed upon me, has given me the strength to continue this challenging and inspiring journey until the end, which has made me who I am and brought me to where I am today. It is with a great relief that I finally finished this research with a lot of gratitude to those who have helped me in any way to reach this point. Without their help, this thesis can never make its way. For that, I believe they deserve a special mention.

First and foremost, my heartiest gratitude goes to my former main supervisor who is now happily retired but still willing to supervise me until the end of my PhD journey, Professor Dr Nor Laila Mohd Noor, for her guidance, courage, support and friendship that were offered to me throughout the completion of my research. I am greatly indebted for all the hours she spent in guiding me, reading and commenting my writings. I appreciate her experience and expertise that she has shared with me which has helped me in completing this research. My special thank also goes to my current main supervisor, Dr Nurulhuda Noordin who has been willing to share her experience and valuable advices that have helped me throughout this journey. Besides that, I am also thankful to my co-supervisor, Professor Madya Dr Syed Ahmad Sheikh Aljunid for his insightful comments especially on the analysis and writing parts. I also want to extend my gratitude and thankfulness to the research assistant Miss Umi Rahimah, the G2NE developer Raja Liana and the nutritionists Nashrah and Adilah who gave their full cooperation in doing the tasks entrusted to them.

To my dear beloved and understanding husband, Nik Shafrizan, a very special thank you for being supportive in every winding path of the journey. Thanks for always being there for me when I need you the most. To my dear daughters, Ariana and Arissa, thanks for being the source of positive spirit and happy pills during the good and bad times. To all my friends, who always shared their ideas, knowledge, and experiences in this journey. Last but not least, I would like to dedicate this work to all my family members, my father Hj Daud Hj Yusoff, Nik Danya Nik Abdul Rahman (in law), my beloved mother Noryati Rutain, Wan Sharipah Wan Ahmad (in law), my dear siblings, and siblings (in law). Thank you for teaching me how beautiful life is, to love and being loved.

TABLE OF CONTENTS

Page

CONFIRMATION BY PANEL OF EXAMINERS **ii**

AUTHOR'S DECLARATION **iii**

ABSTRACT **iv**

ACKNOWLEDGEMENT **v**

TABLE OF CONTENTS **vi**

LIST OF TABLES **xi**

LIST OF FIGURES **xiv**

CHAPTER ONE INTRODUCTION **1**

1.1 Research Background 1

1.2 Preliminary Studies 4

1.2.1 First Preliminary Study (Preliminary Study with HE1) 5

1.2.2 Second Preliminary Study (Preliminary Study with HE2) 7

1.2.3 Third Preliminary Study (Preliminary Study with NE1) 11

1.3 Research Motivation 12

1.3.1 The limited health forecasting based on trends 13

1.3.2 High cost of conducting population-based survey to measure obesity prevalence 13

1.3.3 Assumption of constant increases rate of obesity 13

1.3.4 No nutrition variable is considered in current prediction method. 14

1.4 Research Aim 14

1.5 Problem Statement 15

1.6 Research Claim 16

1.7 Research Questions 16

1.8 Research Objectives 17

1.9 Research Approach 17

1.10 Scope of the Research 21